GLUCURONYL C5-EPIMERASE, DNA ENCODING THE SAME AND USES THEREOF

Inventors: Markku Jalkanen
Kamel El Darwish
Ulf Lindahl
Jin-Ping Li

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of provisional U.S. Appln. No. 60/304,180, which was converted from U.S. Appln. No. 09/732,026 filed December 8, 2000, the contents of which are incorporated herein by reference.

STATEMENT REGARDING FEDERALLY-SPONSORED RESEARCH AND DEVELOPMENT

Not applicable.

Field of the Invention

[0001] The invention is in the field of recombinant proteins, and especially, glucuronyl C5-epimerases and the use of the same for the modification of glucosaminoglycans.

BACKGROUND OF THE INVENTION

[0002] Glucuronyl C5-epimerase (herein, "C5-epimerase") catalyzes the conversion of D-glucuronic acid (GlcA) to L-iduronic acid (IdoA) in the second polymer modification step of heparin/heparan sulfate (HS) synthesis. The epimerase involved in heparin/HS synthesis has an absolute requirement for N-sulfate at the nonreducing side of the target HexA, the formation of which is catalyzed by a N-Deacetylase-N-sulfotransferase (NDST) in the first (preceding)

step of biosynthetic polymer modification. Also, the epimerase is inhibited by O-sulfate groups near its site of action, so O-sulfation steps later in the heparin biosynthetic pathway inhibit epimerization or back-epimerization. The reaction involves reversible abstraction and readdition of a proton at C5 of the target hexuronic acid, via carbanion intermediate, and is believed to involve two polyprotic basic amino acids (esp. Lys).

[0003] The C5-epimerase, like other enzymes involved in heparin/HS biosynthesis, appears to be membrane bound or associated in the Golgi. Interestingly, solubilized epimerase catalyzes both (reversible) reactions, but no back-epimerization is detectable from microsomal fractions. C5-epimerase active protein was first purified and characterized from liver (Campbell *et al.*, *J. Biol. Chem. 269*:26953-26958 (1994)).

[0004] Campbell, P. *et al.*, reported the purification of the D-glucuronyl C5-epimerase from bovine liver (Campbell *et al.*, *J. Biol. Chem. 269*: 26953 -26958 (1994)), and several DNA sequences have also been reported. While the predicted size of the bovine C5-epimerase from genomic and cDNA sequences is 70.1 KD (618 amino acids) (discussed below), the most purified native preparate extracted as above contained predominant species of 52 and 20 kDa, indicating that proteolytic cleavage (processing) may have occurred. Detection of activity in larger MW (200 kDa) fractions from size-exclusion chromatography indicated that aggregation or oligomerization may occur. The enzyme has a broad pH range (6.5-7.5) of activity, having an optimum 7.4. The enzyme does not have a metal ion or other cofactor requirement. Kinetic studies unexpectedly revealed that the K_m increases with increasing enzyme concentration, probably relating to polymeric substrate and stearic hindrance, and/or oligomerization of the epimerase molecules.

[0005] Recently, Lindahl, U. and Li, J-P., WO98/48006, purified the 52 kDa C5-epimerase from bovine liver and obtained a partial amino acid sequence. Primers were made against an internal sequence and used to amplify a sequence from a bovine liver cDNA preparation. The bovine liver sequence was used to screen

a bovine lung cDNA library. A sequence having an open reading frame of 444 amino acids was found, which corresponded to a polypeptide of 49.9 kDa. It was stated that the enzyme previously isolated from bovine liver was a truncated form of the native protein. Total RNA from bovine liver, lung and mouse mastocytoma were analyzed by hybridization to the bovine lung epimerase cDNA clone. Both bovine liver and bovine lung gave identical results, with a dominant transcript of about 9 kb and a weak 5 kb band. The mouse mastocytoma RNA only showed the transcript at about 5 kb.

[0006] The report of the cloning of a cDNA encoding a C5-epimerase from bovine lung also appeared in Li *et al. J. Biol. Chem. 272*: 28158-28163 (1997). Li *et al.* cloned and expressed the bovine lung epimerase in a baculovirus/insect cell system, which first assigned activity to a cloned (recombinant) sequence. The active recombinant protein was not purified for definitive assignment.

[0007] C5-epimerase cDNA sequences from Drosophila (GenBank Accession Number AAF57373), *C. elegans* (GenBank Accession Number P46555)and Methanococcus (GenBank Accession Number U67555) have been reported.

[0008] The enzymatic activity of the recombinant bovine epimerase reported by Lindahl *et al.* was relatively low. However, attempts to express the bovine lung C5-epimerase, the sole cloned mammalian epimerase, in systems that might yield a better production failed. Expression in mammalian cells, *Saccharomyces cerevisiae*, and *E. coli* have been attempted. To date, there have been no reports of the successful production of a soluble, active C5-epimerase. Therefore, it has not been possible to expand the early baculovirus cell system results into other recombinant systems or to use conventional expression methods such as mammalian, yeast and bacterial systems for expression of this enzyme.

[0009] Thus, there remains a need in the art for a highly active C5-epimerase, and for methods for production of larger amounts of the same.

SUMMARY OF THE INVENTION

[0010] Recognizing that problems of an undefined nature exist with expressing recombinant epimerases of mammalian origin, and cognizant of the need for a useful method for expressing and producing useful amounts of the C5-epimerase, the inventors investigated recombinant C5-epimerase production methods. The studies culminated in the discovery of a novel mouse gene, and the mouse C5-epimerase protein encoded therein. The mouse C5-epimerase of the invention is unique, *inter alia*, in that it contains additional sequences at its N-terminus in comparison to the C5-epimerase protein sequences known in the art. It has been unexpectedly discovered that the fusion of the mouse C5-epimerase's N-terminal fragment, or shortened versions thereof, to the N-terminus of other C5-epimerases, greatly enhances the activity of those other recombinant C5-epimerase activity by orders of magnitude. Thus the mouse N-terminus extension can be used to facilitate expression of sequences that are operably linked to it, and especially, expression of native (murine liver) and heterologous (both non-murine and murine non-hepatic) forms of C5-epimerases in recombinant systems.

[0011] Accordingly, in a first embodiment, the invention is directed to purified and/or isolated polynucleotides encoding a mouse (murine) liver C5-epimerase, and recombinant vectors and hosts for the maintenance and expression of the same.

[0012] In a further embodiment, the invention is directed to the purified and/or isolated mouse liver C5-epimerase protein encoded by such polynucleotides, or preparations containing the same.

[0013] In a further embodiment, the invention is directed to methods of producing the mouse C5-epimerase using such polynucleotides and the recombinant vectors and hosts of the invention to express the same.

[0014] In a further embodiment, the invention is directed to polynucleotides, especially purified and/or isolated polynucleotides, encoding a fusion protein,

such fusion protein containing the N-terminal sequence of the mouse C5-epimerase, operably linked in-frame to the amino acid sequence of a desired protein, and especially, a heterologous C5-epimerase sequence, and vectors and hosts for the maintenance and expression of such polynucleotides.

[0015] In a further embodiment, the invention is directed to the purified and/or isolated C5-epimerase fusion protein encoded by such polynucleotides.

[0016] In a further embodiment, the invention is directed to methods of producing a desired protein, by operably linking a polynucleotide that encodes the mouse C5-epimerase, or its N-terminal sequence, to a polynucleotide that encodes such desired protein of interest, and expressing the same in a recombinant host of the invention.

[0017] In a further embodiment, the invention is directed to polynucleotide sequences and vectors that provide polynucleotides encoding the N-terminal fragment polynucleotide sequence of mouse C5-epimerase, such polynucleotides and vectors having desired restriction sites at the 3′-terminus of the fragment for insertion (linkage) of a desired sequence thereto, especially, a sequence that encodes a protein of interest, and most especially, another epimerase sequence.

[0018] In a further embodiment, the invention is directed to methods of using the N-terminal sequence of mouse C5-epimerase for the expression of native and heterologous sequences linked thereto.

BRIEF DESCRIPTION OF THE FIGURES

[0019] **Figure 1**. DNA sequence of a fusion protein having the sequence of the bovine C5-epimerase (non-bold) and the N-terminus of the mouse C5-epimerase (bold). The open reading frame (ORF) showing the polypeptide coding sequence is underlined.

[0020] **Figure 2.** The complete DNA sequence of mouse C5-epimerase.

[0021] **Figure 3**. The complete amino acid sequence of mouse C5-epimerase.

[0022] **Figure 4.** Alignment analysis of mouse C5-epimerase to other sequences showing regions of homology. The scores are shown on the top line and are listed in the column after the source of the sequence. The sequences are taken from the following sources: line 2: mouse liver; line 3: bovine lung; line 4: human EST; line 5: Drosophila; line 6: C. elegans; line 7: Methanococcus.

[0023] **Figure 5.** Diagrammatic representation of the domain structure of the mouse C5-epimerase. Solid rectangular box at the N-terminus: signal sequence (highly hydrophobic transmembrane (TM) sequence); hatched rectangular boxes: hydrophobic transmembrane (TM) or buried sequences; solid rectangular boxes within the peptide: conserved peptide sequences having greater than 50% similarity to the *C. elegans* 71.9 KD hypothetical protein.

[0024] **Figure 6A-6B.** Figure 6A: Diagrammatic representations of the products of the tagged recombinant (bovine) C5-epimerase constructions. i: First active tagged recombinant (bovine) C5-epimerase construct. The specific activity was $5x10^5$ cpm/mg/h). ii: The most active recombinant (full mouse) C5 construction. The specific activity was $2x10^9$ cpm/mg/h. iii: Chimeric construct having both mouse and bovine sequences. The activity was 87% of the activity of the full-length mouse sequence. iv: Truncated mouse construct. The activity is the same as the bovine construct in "i". Figure 6B: sequence and domain information of the tag that preceded each of the recombinant constructs in Figure 6A.

[0025] **Figure 7.** Activity assay results of mouse C5-epimerase (mC5).

[0026] **Figure 8.** Western blot stained with anti-FLAG. Lane 1: molecular weight standards (New England Biolabs' Broad Range, prestained). Lane 2: mouse C5-epimerase (mC5) sample.

[0027] **Figure 9.** Western blot of the proteins in the medium from stable insect cell lines of clones containing different tagged recombinant C5-epimerases stained with anti-FLAG antibody.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0028] A mouse liver gene encoding C5-epimerase was cloned. The nucleotide sequence is shown in Figure 2. The amino acid sequence of the mouse liver C5-epimerase protein was found to be 618 amino acids long (Figure 3), with a molecular weight of 71,180.1 daltons (71.18 kDa). The mouse C5-epimerase has an isoelectric point of 8.25 and a net charge at pH 7 of +4.01.

[0029] The amino acid sequence of the mouse liver C5-epimerase sequence, without any N-terminal extension is homologous (>96% amino acid identity) to the bovine C5-epimerase sequence. However, sequence analysis revealed that the N-terminus of the enzyme that is encoded by the mouse genomic sequence contained an additional 154 amino acids (amino acid) that were "missing" from the cloned bovine sequence.

[0030] The mouse coding sequence displayed >95% peptide identity to the corresponding bovine and human (expressed sequence tag from brain cDNA library) sequences, >50% similarity to a hypothetical 71.9 kDa protein from expressed sequence of *C. elegans*, and 38% similarity to a protein from an expressed sequence of *Methanococcus sp.*

[0031] The predicted transmembrane topology (hydrophobicity plot) of the mouse C5-epimerase enzyme resembles that of NDST. These and other observations (e.g., speed of heparin synthesis) indicated that the C5-epimerase and other enzymes of heparin biosynthesis are likely associated in a complex *in vivo*.

[0032] The recombinant mouse C5-epimerase, as expressed and secreted by an insect cell signal, from the baculovirus insect cell system, is most stable in medium at 4°C. Purification of the recombinant C5-epimerase may include, but is not limited to, such processes as cation exchange or affinity chromatography. For example, the recombinant protein may be engineered such that the protein contains a FLAG-tag or His-tag that occurs at either end of the recombinant protein. As one of ordinary skill in the art will appreciate, in such instances, the

recombinant protein may be purified using commercially available resins which utilize, for example, anti-FLAG monoclonal antibodies to capture the recombinant protein comprising the FLAG epitome.

[0033] The enzyme is most rapidly assayed by biphasic extraction of tritium released from C5-labeled substrate into an organic scintillation cocktail, and counting, though ultimate confirmation of activity is by NMR analysis of converted product as described in the examples.

[0034] The native mouse liver enzyme has a specific activity of 5-10 X 10^9 cpm/mg/h, while that of the recombinant form of the mouse enzyme was about 2 X 10^9 cpm/mg/h. By comparison, the recombinant bovine enzyme has a specific activity of about 0.5-1.0 X 10^6 cpm/mg/h. Therefore, the recombinant mouse enzyme is an especially active C5-epimerase.

[0035] Unexpectedly, it was found that the 154 amino acid (amino acid) N-terminus of the mouse C5-epimerase, and especially certain fragments thereof, have the ability of being greatly able to enhance the enzymatic C5-epimerase activity of other C5-epimerases when fused in-frame to the N-terminus of the same. This additional 154 amino acid (amino acid) fragment appears to have at least three features that are desirable for the recombinant expression and secretion of an active C5-epimerase. First, it includes a sequence that is thought to function as a signal sequence comprised of the first 33-34 residues (amino acids 1-33 or 1-34 of Figure 3). Second, it provides additional cysteine residues that are amenable for the formation of disulfide bonds and for the stabilization of secondary protein structure. Third, it provides an amidation site that is consistent with a site useful for posttranslational proteolytic processing.

[0036] Fragments of the 154 amino acid sequence that lack the signal sequence still possess the ability to enhance the activity of heterologous epimerases, and especially C5-epimerases, to which they are operably linked. For example, as shown in the examples, a fusion protein that contains amino acids 34-154 directly linked, in-frame to the N-terminus of the bovine C5-epimerase enhanced the activity of the bovine C5-epimerase over 100-fold.

Nucleic Acid Molecules

[0037] The present invention provides isolated nucleic acid molecules, comprising:

(1) a polynucleotide encoding the mouse liver C5-epimerase polypeptide having the amino acid sequence shown in Figure 3.

(2) a polynucleotide encoding useful fragments of the mouse liver C5-epimerase polypeptide having the amino acid sequence shown in Figure 3, such useful fragments including but not limited to (a) fragments that provide the signal sequence of amino acids 1-33 or 1-34; (b) fragments that provide the mature mouse liver C5-epimerase protein sequence, and especially amino acids 33-618 or 34-618, and (c) fragments that provide the sequence of the activity-stimulating N-terminus fragment having amino acids 1-154, and including fragments thereof such as amino acids 33-154 or 34-154 that possess the ability to enhance the activity of other C5-epimerases to which they are operably linked;

[0038] Unless otherwise indicated, all nucleotide sequences determined by sequencing a DNA molecule herein were determined as described in the examples, and all amino acid sequences of polypeptides encoded by DNA molecules determined herein were predicted by translation of a DNA sequence determined as above. Therefore, as is known in the art for any DNA sequence determined by this approach, any nucleotide sequence determined herein may contain some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least about 95% to at least about 99.9% identical to the actual nucleotide sequence of the sequenced DNA molecule. The actual sequence can be more precisely determined by other approaches including manual DNA sequencing methods which are well known in the art. As is also known in the art, a single insertion or deletion in a determined nucleotide sequence compared to the actual sequence will cause a frame shift in translation of the nucleotide sequence such that the predicted amino

acid sequence encoded by a determined nucleotide sequence will be completely different from the amino acid sequence actually encoded by the sequenced DNA molecule, beginning at the point of such an insertion or deletion.

[0039] By "nucleotide sequence" of a nucleic acid molecule or polynucleotide is intended, for a DNA molecule or polynucleotide, a sequence of deoxyribonucleotides, and for an RNA molecule or polynucleotide, the corresponding sequence of ribonucleotides (A, G, C and U), where each thymidine deoxyribonucleotide (T) in the specified deoxyribonucleotide sequence is replaced by the ribonucleotide uridine (U).

[0040] Using the information provided herein, such as the nucleotide sequence set out in Figures and sequence listing, a nucleic acid molecule of the present invention encoding a C5-epimerase polypeptide, or a chimeric construct of the same, may be obtained using standard cloning and screening procedures, such as those for cloning cDNAs using mRNA as starting material. Illustrative of the invention, the C5-epimerase nucleic acid molecule described in Figures 2 and 3 was discovered in a cDNA library derived from murine hepatic (liver) tissue.

[0041] The determined nucleotide sequence of the C5-epimerase DNA of Figure 2 contains an open reading frame encoding a protein of about 618 amino acid residues, with an initiation codon at nucleotide position 1 of the nucleotide sequences in Figure 2.

[0042] As one of ordinary skill would appreciate, due to the possibility of sequencing errors discussed above, the actual complete C5-epimerase polypeptide encoded by the sequence of Figure 2, which comprise about 618 amino acids as shown in Figure 3, may be somewhat longer or shorter. In any event, as discussed further below, the invention further provides polypeptides having various residues deleted from the N-terminus or the C-terminus of the complete polypeptide, including polypeptides lacking one or more amino acids from the N-terminus or C-terminus of the extracellular domain described herein.

[0043] The nucleic acid molecules of the invention include those that encode the C5-epimerase a signal sequence, as shown in Figure 3, which is amino acids 1-33

or amino acids 1-34 of the amino acid sequence shown in Figure 3. Such molecules can be operably linked in-frame to any desired nucleotide sequence, especially one that encodes a protein of interest that it is desired to secrete from a host in which the C5-epimerase signal sequence is capable of secreting.

[0044] Additionally, the nucleic acid molecules of the invention include those that encode the mouse liver C5-epimerase's "heterologous activity enhancing" sequence which is amino acids 1-154, or at least 30 amino acids thereof, as shown in Figure 3. What is meant by the term "heterologous" nucleic acid is well known to one of ordinary skill in the art as being derived from the nucleic acid of a different species. Preferably, such nucleic acid molecules encode amino acids 1-154, 33-154 or 34-154 as shown in Figure 3, plus or minus 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acids from either or both ends. A nucleic acid encoding such a polypeptide can be operably linked, in frame, to the coding sequence for another epimerase, especially another C5-epimerases, with the result that a fusion protein is encoded by the nucleic acid construct. In a preferred embodiment, the heterologous activity enhancing sequences are expressed at the N-terminus of the fusion protein and are linked to the N-terminus of another protein whose activity is enhanced by the presence of the mouse sequence, most especially a non-mouse C5-epimerase or an isozyme of the mouse C5-epimerase.

[0045] As indicated, nucleic acid molecules of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance, cDNA and genomic DNA obtained by cloning or produced synthetically. The DNA may be double-stranded or single-stranded. Single-stranded DNA or RNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

[0046] By "isolated" nucleic acid molecule(s) is intended a nucleic acid molecule, DNA, or RNA, which has been removed from its native environment. For example, recombinant DNA molecules contained in a vector are considered isolated for the purposes of the present invention. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous

host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

[0047] Isolated nucleic acid molecules of the present invention include DNA molecules comprising an open reading frame (ORF) that encodes a C5-epimerase protein or fusion protein of the invention. DNA molecules comprising the coding sequence for the C5-epimerase protein as shown in Figure 2, or desired fragment thereof; and DNA molecules which comprise a sequence substantially different from those described above, but which, due to the degeneracy of the genetic code, still encode the C5-epimerase protein amino acid sequence as shown in Figure 3. Of course, the genetic code is well known in the art. Thus, it would be routine for one skilled in the art to generate such degenerate variants. In a further embodiment, nucleic acid molecules are provided that encode the C5-epimerase polypeptide as above, but lacking the N-terminal methionine, or the signal sequence encoded by amino acids 1-33 or 1-34 as shown on Figure 3, or having the coding sequence of a different (heterologous) signal sequence operably linked thereto.

[0048] The invention further provides not only the nucleic acid molecules described above but also nucleic acid molecules having sequences complementary to the above sequences. Such isolated molecules, particularly DNA molecules, are useful as probes for gene mapping, by *in situ* hybridization with chromosomes, and for detecting expression of the C5-epimerase gene in various species and tissues, for instance, by Northern blot analysis.

[0049] The present invention is further directed to fragments of the isolated nucleic acid molecules described herein that retain a desired property or that encode a polypeptide that retains a desired property or activity. By a fragment of an isolated nucleic acid molecule as described above is intended fragments at least about 15 nucleotides (nucleotide), and more preferably at least about 20 nucleotide, still more preferably at least about 30 nucleotide, and even more

preferably, at least about 40 nucleotide in length which are useful as diagnostic probes and primers as discussed herein, or to provide a desired motif or domains to a fusion protein construct. Of course, larger fragments 50-300 nucleotide, or even 600 nucleotide in length are also useful according to the present invention as are fragments corresponding to most, if not all, of the nucleotide sequence of the DNA shown in Figure 2 or encoding the amino acid sequence of Figure 3. By a fragment at least 20 nucleotide in length when compared to that of Figure 2, for example, is intended fragments which include 20 or more contiguous bases from the nucleotide sequence of the nucleotide sequence as shown in Figure 2.

[0050] In particular, the invention provides polynucleotides having a nucleotide sequence representing the portion of that shown in Figure 2 or encoding the amino acid sequence shown in Figure 3. Also contemplated are polynucleotides encoding C5-epimerase polypeptides which lack an amino terminal methionine. Polypeptides encoded by such polynucleotides are also provided, such polypeptides comprising an amino acid sequence at positions 2 to 618 of the amino acid sequence shown on Figure 3, but lacking an amino terminal methionine.

[0051] In another aspect, the invention provides an isolated nucleic acid molecule comprising a polynucleotide which hybridizes under stringent hybridization conditions to a portion or preferably all of the polynucleotide in a nucleic acid molecule of the invention described above. By a portion could be any desired portion, for example, the polynucleotide of Figure 2 that encode amino acids 1-154 or 33-154 or 34-154. By "stringent hybridization conditions" is intended overnight incubation at 42° C in a solution comprising: 50% formamide, 5x SSC (750 mM NaCl, 75 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 μg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65° C.

[0052] By a polynucleotide which hybridizes to a "portion" of a polynucleotide is intended a polynucleotide (either DNA or RNA) hybridizing to at least about 15 nucleotides (nucleotide), and more preferably at least about 20 nucleotide, still

more preferably at least about 30 nucleotide, and even more preferably about 30-70 (e.g., 50) nucleotide of the reference polynucleotide. These are useful as diagnostic probes and primers as discussed above and in more detail below.

[0053] By a portion of a polynucleotide of "at least 20 nucleotide in length," for example, is intended 20 or more contiguous nucleotides from the nucleotide sequence of the reference polynucleotide (e.g., the nucleotide sequence as shown in Figure 2). Of course, a polynucleotide which hybridizes only to a poly A sequence, or to a complementary stretch of T (or U) residues, would not be included in a polynucleotide of the invention used to hybridize to a portion of a nucleic acid of the invention, since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

[0054] As indicated, nucleic acid molecules of the present invention which encode a C5-epimerase polypeptide may include, but are not limited to the coding sequence for the polypeptide, by itself (also called the mature C5-epimerase when it lacks the secretion signal); the coding sequence for the polypeptide and additional sequences, such as those encoding a leader or secretary sequence, such as a pre-, or pro- or prepro- protein sequence; the coding sequence of the polypeptide, with or without the aforementioned additional coding sequences, together with additional, non-coding sequences, including for example, but not limited to introns and non-coding 5′ and 3′ sequences, such as the transcribed, non-translated sequences that play a role in transcription, mRNA processing - including splicing and polyadenylation signals, for example - ribosome binding and stability of mRNA; additional coding sequence which codes for additional amino acids, such as those which provide additional functionalities. Thus, for instance, the polypeptide may be fused to a marker sequence, such as a peptide, which facilitates purification of the fused (marker containing) polypeptide. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (Qiagen, Inc.), among others, many of which are commercially available. As

described in Gentz *et al., Proc. Natl. Acad. Sci. USA 86*: 821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. The "HA" tag is another peptide useful for purification which corresponds to an epitope derived from the influenza hemagglutinin protein, which has been described by Wilson *et al., Cell 37*:767-778(1984).

Variant and Mutant Polynucleotides

[0055] The present invention further relates to variants of the nucleic acid molecules of the present invention, which encode portions, analogs, or derivatives of the C5-epimerase. Variants may occur naturally, such as a natural allelic variant. By an "allelic variant" is intended one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. *Genes II*, Lewin, B., ed., John Wiley & Sons, New York (1985). Non-naturally occurring variants may be produced using art-known mutagenesis techniques.

[0056] Such variants include those produced by nucleotide substitutions, deletions or additions. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding regions, non-coding regions, or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the C5-epimerase polypeptide or portions thereof. Also especially preferred in this regard are conservative substitutions.

[0057] Further embodiments of the invention include an isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence encoding a polypeptide, the amino acid sequence of which is at least 80% identical to, and more preferably at least 90%, 95%, 96%, 97%, 98% or 99% identical to, a reference amino acid sequence selected from the group consisting of: (a) amino acids 1 to 118 of Figure 3; (b) amino acids 1 to 119 of Figure 3; (c) amino acids 1 to 120 of Figure 3; (d) amino acids 1 to 121 of Figure 3; (e) amino acids 119 to

618 of Figure 3; (f) amino acids 120 to 618 of Figure 3; (g) amino acids 121 to 618 of Figure 3; (h) amino acids 122 to 618 of Figure 3; (i) amino acids 34 to 147 of Figure 3; (j) amino acids 35 to 154 of Figure 3; (k) amino acids 34 to 154 of Figure 3; and (l) amino acids 1 to 154 of Figure 3; (m) the entire amino acid sequence shown on Figure 3.

[0058] Further embodiments of the invention include isolated nucleic acid molecules that comprise a polynucleotide which hybridizes under stringent hybridization conditions to a polynucleotide in (a), (b), (c), (d), (e), (f), (g), (h), (i), (j), (k), (l), (m), (n), above. This polynucleotide which hybridizes does not hybridize under stringent hybridization conditions to a polynucleotide having a nucleotide sequence consisting of only A residues or of only T residues.

[0059] By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence encoding a C5-epimerase polypeptide is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the C5-epimerase polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5′ or 3′ terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence.

[0060] As a practical matter, whether any particular nucleic acid molecule is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the nucleotide sequence shown in Figure 2 can be determined conventionally using known computer programs such as the Bestfit program (Wisconsin Sequence Analysis

Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711). Bestfit uses the local homology algorithm of Smith and Waterman, *Advances in Applied Mathematics 2*:482-489 (1981), to find the best segment of homology between two sequences. When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence according to the present invention, the parameters are set, of course, such that the percentage of identity is calculated over the full length of the reference nucleotide sequence and that gaps in homology of up to 5% of the total number of nucleotides in the reference sequence are allowed.

[0061] The present application is directed to nucleic acid molecules at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleic acid sequence shown in Figure 2, irrespective of whether it encode a polypeptide having C5-epimerase activity. This is because even where a particular nucleic acid molecule does not encode a polypeptide having C5-epimerase activity, one of skill in the art would still know how to use the nucleic acid molecule, for instance, as a hybridization probe or a polymerase chain reaction (PCR) primer. Uses of the nucleic acid molecules of the present invention that do not encode a polypeptide having C5-epimerase activity include, inter alia: (1) isolating a C5-epimerase gene or allelic variants thereof in a cDNA library; (2) in situ hybridization (e.g., "FISH") to metaphase chromosomal spreads to provide precise chromosomal location of the C5-epimerase gene, as described in Verma *et al.*, Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York (1988); and Northern Blot analysis for detecting C5-epimerase mRNA expression in specific tissues.

[0062] Preferred, however, are nucleic acid molecules having sequences at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleic acid sequence shown in Figure 2 which does, in fact, encode a polypeptide having C5-epimerase activity. By "a polypeptide having C5-epimerase activity" is intended polypeptides exhibiting activity similar, but not necessarily identical, to an activity of the C5-epimerase of the invention (either the full length protein or

preferably the identified amino acid fragment containing amino acids 33-618 or 34-618), as measured in a particular biological assay.

[0063] Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of the nucleic acid molecules having a sequence at least 80%, 90%, 95%, 96%, 97%, 98%, or 99% identical to the nucleic acid sequence of a deposited cDNA or the nucleic acid sequence shown in Figure 2 will encode a polypeptide "having C5-epimerase protein activity." In fact, since degenerate variants of these nucleotide sequences all encode the same polypeptide, this will be clear to the skilled artisan even without performing the above described comparison assay. It will be further recognized in the art that, for such nucleic acid molecules that are not degenerate variants, a reasonable number will also encode a polypeptide having C5-epimerase protein activity. This is because the skilled artisan is fully aware of amino acid substitutions that are either less likely or not likely to significantly effect protein function (e.g., replacing one aliphatic amino acid with a second aliphatic amino acid), as further described below.

Vectors and Host Cells

[0064] The present invention also relates to vectors which include the isolated DNA molecules of the present invention, host cells which are genetically engineered with the recombinant vectors of the invention and the production of C5-epimerase polypeptides or fragments thereof by recombinant techniques.

[0065] The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged *in vitro* using an appropriate packaging cell line and then transduced into host cells.

[0066] The DNA insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the *E. coli lac, trp* and *tac* promoters, the

SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination and, in the transcribed region, a ribosome binding site for translation. The coding portion of the mature transcripts expressed by the constructs will preferably include a translation initiating at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

[0067] As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase or neomycin resistance for eukaryotic cell culture and tetracycline or ampicillin resistance genes for culturing in *E. coli* and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, *Corynebacterium, Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as *Aspergillus, Aspergillus niger,* or *Trichoderma,* or yeast cells such as *Saccharomyces, Saccharomyces cerevisiae;* insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS and Bowes melanoma cells; and plant cells. Preferred hosts includes insect cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

[0068] Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from Qiagen; pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Viral vectors include, but are not limited to retroviral vectors, pox virus vectors, including vaccinia virus and adenoviral vectors. Other suitable vectors will be readily apparent to the skilled artisan.

[0069] Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic

lipid-mediated transfection, electroporation, transduction, infection or other methods. Such methods are described in many standard laboratory manuals, such as Davis *et al., Basic Methods In Molecular Biology* (1986).

Polypeptides and Fragments

[0070] The invention further provides an isolated or purified C5-epimerase polypeptide having the amino acid sequences encoded by the amino acid sequences in Figure 3, or a peptide or polypeptide comprising a portion of the above polypeptide, especially as described above and encoded by a nucleic acid molecule described above.

[0071] The invention further provides fusion proteins containing a functional portion of the N-terminus of the mouse C5-epimerase, fused at its C-terminus to the N-terminus of a protein of interest, such as, for example, the signal sequence of amino acids 1-33 or 1-34 as shown on Figure 3, or the activity enhancing sequence of amino acids 1-154, 33-154 or 34-154 as shown on Figure 3. In one embodiment, the protein of interest is fused to a portion of the N-terminus that contains from 30 to 154 amino acids of the N-terminus of mouse C5-epimerase of Figure 3, and especially amino acids 33-154 or 34-154. In another preferred embodiment, the protein of interest is fused to a functional portion of the N-terminus that contains residues 33-154 of the sequence shown on Figure 3. In a highly preferred embodiment, the protein of interest is fused to a functional portion of the N-terminus that contains the secretion signal of amino acids 1-33 or 1-34 as shown in the sequence on Figure 3.

[0072] The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals but also additional heterologous functional regions. What is meant by the term "heterologous" polypeptide is well known to one of ordinary skill in the art as being derived from different species. Thus, for instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the

polypeptide to improve stability and persistence in the host cell, during purification or during subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to polypeptides to engender secretion or excretion, to improve stability and to facilitate purification, among others, are familiar and routine techniques in the art.

[0073] The C5-epimerase or fusion protein containing a fragment thereof can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography.

[0074] Polypeptides of the present invention include naturally purified products, products of chemical synthetic procedures, and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. The polypeptide of the instant invention may also include a modification of a histidine or poly histidine added to the termini for protein purification procedures.

[0075] C5-epimerase polynucleotides and polypeptides may be used in accordance with the present invention for a variety of applications, particularly those that make use of the chemical and biological properties of C5-epimerase. Specifically, the recombinant epimerases of the present invention may be used to produce heparin and/or heparan sulfate, which may be useful as anticoagulants, on a larger scale. Also, the epimerases of the present invention may be useful in an experimental setting for studying the effects of extracellular matrix molecules

such as heparin and heparan sulfate on such processes as embryology, angiogenesis and tumor progression. For example, the enzyme can modulate the ratio of D-glucuronic acid/L-iduronic acid residues in heparin or heparan sulfate. L-iduronic acid residues, due to their unique conformational properties, are believed to promote interactions of polysaccharides with proteins. Additionally, the epimerases of the current invention may also be used modify industrially useful sugars which may be used as a stabilizer or gelling agent in some foods.

Variant and Mutant Polypeptides

[0076] To improve or alter the characteristics of a C5-epimerase polypeptide, protein engineering may be employed. Recombinant DNA technology known to those skilled in the art can be used to create novel mutant proteins or "muteins" including single or multiple amino acid substitutions, deletions, additions or fusion proteins. Such modified polypeptides can show, e.g., enhanced activity or increased stability. In addition, they may be purified in higher yields and show better solubility than the corresponding natural polypeptide, at least under certain purification and storage conditions.

N-Terminal and C-Terminal Deletion Mutants

[0077] For instance, for many proteins, including the extracellular domain of a membrane associated protein or the mature form(s) of a secreted protein, it is known in the art that one or more amino acids may be deleted from the N-terminus or C-terminus without substantial loss of biological function. For instance, Ron *et al., J. Biol. Chem., 268*:2984-2988 (1993), reported modified KGF proteins that had heparin binding activity even if 3, 8, or 27 amino-terminal amino acid residues were missing.

[0078] However, even if deletion of one or more amino acids from the N-terminus of a protein results in modification or loss of one or more biological

functions of the protein, other biological activities may still be retained. Thus, the ability of the shortened protein to induce and/or bind to antibodies which recognize the complete or portion of the C5-epimerase protein generally will be retained when less than the majority of the residues of the complete protein or extracellular domain are removed from the N-terminus. Whether a particular polypeptide lacking N-terminal residues of a complete protein retains such immunologic activities can readily be determined by routine methods described herein and otherwise known in the art.

[0079] Accordingly, the present invention further provides polypeptides having one or more residues deleted from the amino terminus of the amino acid sequence shown in Figure 3.

[0080] However, even if deletion of one or more amino acids from the C-terminus of a protein results in modification or loss of one or more biological functions of the protein, other biological activities may still be retained. Thus, the ability of the shortened protein to induce and/or bind to antibodies which recognize the complete or mature form of the protein generally will be retained when less than the majority of the residues of the complete or mature form protein are removed from the C-terminus. Whether a particular polypeptide lacking C-terminal residues of a complete protein retains such immunologic activities can readily be determined by routine methods described herein and otherwise known in the art.

[0081] The invention also provides polypeptides having one or more amino acids deleted from both the amino and the carboxyl termini.

Other Mutants

[0082] In addition to terminal deletion forms of the protein discussed above, it will also be recognized by one of ordinary skill in the art that some amino acid sequences of the C5-epimerase polypeptide can be varied without significant effect on the structure or function of the proteins. If such differences in sequence

are contemplated, it should be remembered that there will be critical areas on the protein which determine activity. Thus, the invention further includes variations of the C5-epimerase polypeptide, which show substantial C5-epimerase polypeptide activity or which include regions of C5-epimerase protein such as the protein portions discussed below. Such mutants include deletions, insertions, inversions, repeats, and type substitutions. Guidance concerning which amino acid changes are likely to be phenotypically silent can be found in Bowie, J. U. *et al.*, "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions," *Science 247*:1306-1310 (1990).

[0083] Thus, the fragment, derivative, or analog of the polypeptide of Figure 3 or fusion protein containing the same, may be: (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue(s), and more preferably at least one but less than ten conserved amino acid residue(s)), and such substituted amino acid residue(s) may or may not be one encoded by the genetic code; or (ii) one in which one or more of the amino acid residues includes a substituent group; or (iii) one in which the mature or soluble extracellular polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol).; or (iv) one in which the additional amino acids are fused to a leader or secretory sequence or a sequence which is employed for purification of the mature polypeptide or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

[0084] Thus, the C5-epimerase of the present invention may include one or more amino acid substitutions, deletions or additions, either from natural mutations or human manipulation. As indicated, changes are preferably of a minor nature, such as conservative amino acid substitutions that do not significantly affect the folding or activity of the protein (see Table 1).

TABLE 1. Conservative Amino Acid Substitutions

Aromatic	Phenylalanine Tryptophan Tyrosine
Hydrophobic	Leucine Isoleucine Valine
Polar	Glutamine Asparagine
Basic	Arginine Lysine Histidine
Acidic	Aspartic Acid Glutamic Acid
Small	Alanine Serine Threonine Methionine Glycine

[0085] Amino acids in the C5-epimerase protein of the present invention that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, *Science 244*:1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as receptor binding or *in vitro* proliferative activity.

[0086] Of particular interest are substitutions of charged amino acids with another charged amino acids and with neutral or negatively charged amino acids. The latter results in proteins with reduced positive charge to improve the characteristics of the C5-epimerase protein. The prevention of aggregation is highly desirable. Aggregation of proteins not only results in a loss of activity but can also be problematic when preparing pharmaceutical formulations, because

they can be immunogenic. (Pinckard *et al., Clin Exp. Immunol. 2*:331-340 (1967); Robbins *et al., Diabetes 36*:838-845 (1987); Cleland *et al. Crit. Rev. Therapeutic Drug Carrier Systems 10*:307-377 (1993)).

[0087] The replacement of amino acids can also change the selectivity of binding of a ligand to cell surface receptors. For example, Ostade *et al., Nature 361*:266-268 (1993), describes certain mutations resulting in selective binding of TNF-α to only one of the two known types of TNF receptors. Sites that are critical for ligand-receptor binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al., J. Mol. Biol. 224*:899-904 (1992) and de Vos *et al., Science 255*:306-312 (1992)).

[0088] The polypeptides of the present invention are preferably provided in an isolated form. By "isolated polypeptide" is intended a polypeptide removed from its native environment. The polypeptide produced and/or contained within a recombinant host cell is considered isolated for purposes of the present invention. Also intended as an "isolated polypeptide" are polypeptides that have been purified, partially or substantially, from a recombinant host cell. For example, a recombinantly produced version of the C5-epimerase polypeptide can be substantially purified by the one-step method described in Smith and Johnson, *Gene 67*:31-40 (1988). Preferably, the polypeptide of the invention is purified to a degree sufficient for sequence analysis, or such that it represents 99% of the proteinaceous material in the preparation.

[0089] The present inventors have discovered the mouse C5-epimerase gene and protein, and that the C5-epimerase polypeptide is a 618 residue protein exhibiting an N-terminal 154 amino acid domain, and especially a 33 or 34 amino acid domain containing amino acids 1-33 or 1-34 that is involved in secretion and stabilization of amino acid sequences that are linked to it. Accordingly, this domain, or a functional portion thereof, is useful for expression and secretion of proteins such as the C5-epimerase, or any other protein, especially a protein that associates with the Golgi apparatus or is otherwise associated with heparin or heparan sulfate synthesis.

[0090] The present inventors have also discovered that the N-terminus of the mouse C5-epimerase protein, and especially amino acids 1-154, 33-154 or 34-154, are especially useful to enhance the activity of other enzymes, especially other C5-epimerases. Accordingly, this domain, or a functional portion thereof, is useful for expression and secretion of fusion proteins that include C5-epimerase sequences heterologous to that shown in Figure 3, especially the bovine C5-epimerase.

[0091] The polypeptides of the invention include the C5-epimerase polypeptide and fragments as discussed above, the amino acid sequence of which is at least 80% identical to a sequence selected from the group consisting of: (a) amino acids 1 to 118 of Figure 3; (b) amino acids 1 to 119 of Figure 3; (c) amino acids 1 to 120 of Figure 3; (d) amino acids 1 to 121 of Figure 3; (e) amino acids 119 to 618 of Figure 3; (f) amino acids 120 to 618 of Figure 3; (g) amino acids 121 to 618 of Figure 3; (h) amino acids 122 to 618 of Figure 3; (i) amino acids 34 to 147 of Figure 3; (j) amino acids 35 to 154 of Figure 3; (k) amino acids 34 to 154 of Figure 3; (l) amino acids 1 to 154 of Figure 3; and (m) the complete amino acid sequence as shown in Figure 3.

[0092] The invention includes polypeptides which are at least 80% identical, more preferably at least 90% or 95% identical, still more preferably at least 96%, 97%, 98%, or 99% identical to the polypeptides described above, and also include portions of such polypeptides with at least 30 amino acids and more preferably at least 50 amino acids.

[0093] By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a reference amino acid sequence of a C5-epimerase polypeptide is intended that the amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid of the C5-epimerase polypeptide. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference

sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

[0094] As a practical matter, whether any particular polypeptide is at least 80%, 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequence shown in Figure 3 can be determined conventionally using known computer programs such the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711). When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence according to the present invention, the parameters are set, of course, such that the percentage of identity is calculated over the full length of the reference amino acid sequence and that gaps in homology of up to 5% of the total number of amino acid residues in the reference sequence are allowed.

[0095] The polypeptides of the present invention that possess C5-epimerase activity can be used to provide the same *in vitro,* for example, in developing assays for the same or in standardizing assays for use with more complex systems. The signal sequence of the invention can be used to secrete the homologous C5-epimerase enzyme from eukaryotic recombinant hosts, or to secrete heterologous sequences that are operably linked to the same. The activity enhancing sequence of the invention can be used to enhance the inherent epimerase activity of recombinant preparations of other C5-epimerases, and as such is best provided in the form of a gene encoding a fusion protein for the same.

Antibodies

[0096] C5-epimerase-protein specific antibodies for use in the present invention can be raised against the intact C5-epimerase proteins or an antigenic polypeptide fragment thereof, which may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier, or in liposomes or complexed with PEG to enhance circulatory half-life.

[0097] As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and $F(ab')_2$ fragments) which are capable of specifically binding to a C5-epimerase protein. Fab and $F(ab')_2$ fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding of an intact antibody (Wahl *et al., J. Nucl. Med. 24*:316-325 (1983)). Thus, these fragments are preferred.

[0098] The antibodies of the present invention may be prepared by any of a variety of methods. For example, cells expressing the C5-epimerase protein or an antigenic fragment thereof can be administered to an animal in order to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of C5-epimerase protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

[0099] In the most preferred method, the antibodies of the present invention are monoclonal antibodies. Such monoclonal antibodies can be prepared using hybridoma technology (Köhler *et al., Nature 256*:495 (1975); Köhler *et al., Eur. J. Immunol. 6*:511 (1976); Köhler *et al., Eur. J. Immunol. 6*:292 (1976); Hammerling *et al., in*: *Monoclonal Antibodies and T-Cell Hybridomas*, Elsevier, N.Y., (1981) pp. 563-681). In general, such procedures involve immunizing an animal (preferably a mouse) with a C5-epimerase protein antigen or, more preferably, with a C5-epimerase protein-expressing cell. Suitable cells can be

recognized by their capacity to bind anti-C5-epimerase protein antibody. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 μg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the American Type Culture Collection, Manassas, Virginia. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*, *Gastroenterology 80*:225-232 (1981). The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the desired C5-epimerase antigen.

[0100] Alternatively, additional antibodies capable of binding to the C5-epimerase antigen may be produced in a two-step procedure through the use of anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and that, therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, C5-epimerase-protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the C5-epimerase protein-specific antibody can be blocked by the C5-epimerase protein antigen. Such antibodies comprise anti-idiotypic antibodies to the C5-epimerase protein-specific antibody and can be used to immunize an animal to induce formation of further C5-epimerase protein-specific antibodies.

[0101] It will be appreciated that Fab and $F(ab')_2$ and other fragments of the antibodies of the present invention may be used according to the methods

disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce $F(ab')_2$ fragments). Single chain antibodies, such as light or heavy chain antibodies, are also encompassed by this invention. Alternatively, C5-epimerase protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry. Such antibodies would include, but not be limited to recombinant abs which comprise complementarity determining regions (CDRs) that have differing binding specificities or CDRs which have been modified through the application of recombinant DNA technology or through synthetic chemistry to modify the binding specificity of the antibodies.

[0102] For *in vivo* use of anti-C5-epimerase in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. See, for review, Morrison, *Science 229*:1202 (1985); Oi *et al.*, *BioTechniques 4*:214 (1986); Cabilly *et al.*, U.S. Patent No. 4,816,567; Taniguchi *et al.*, EP 171496; Morrison *et al.*, EP 173494; Neuberger *et al.*, WO 8601533; Robinson *et al.*, WO 8702671; Boulianne *et al.*, *Nature 312*:643 (1984); Neuberger *et al.*, *Nature 314*:268 (1985).

[0103] Bifunctional antibodies are antibodies which have antigen binding domains to different epitopes or derived from different species and are encompassed by the invention. Antibodies with Fc regions derived from species differing from the Fab regions are also envisioned and can be used in immunospecific chromatographic procedures. Also encompassed by this invention are antibodies with attached labels such as fluorescein, Texas Red, rhodamine, peroxidase, gold, magnetic labels, alkaline phosphatase, radioisotopes or chemiluminescent labels .

[0104] Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

EXAMPLES

EXAMPLE 1

Isolation and sequencing of mouse genomic clones

[0105] A mouse genomic library (FIX II, Stratagene) was screened with a DNA probe from a bovine sequence encoding C5-epimerase. The probe was labeled with [α^{32}P]dCTP (NEN Life Science Products). Approximately 2 x 10^6 phages were plated in a 20 x 20 cm plate and duplicate nylon filters were prepared from each plate. High stringency screening was performed with hybridization in 5 x Denhardts, containing 100 µg of salmon sperm DNA/ml at 60°C. The final washes were in 0.1 x SSC (1 x SSC is 150 mM NaCl, 15 mM sodium citrate, pH 7.0 containing 0.1% SDS). Plaques that produced positive signals on both replicas were selected for second and third round screening, and ultimately five positive clones were isolated. It was found that two of the clones have a similar length of about 16 kb, while the other three were relatively shorter, around 10-12 kb. The longest clone (clone 64) was digested with Sac1 and the restriction fragments were cloned into pBlueScript. The second longest clone (clone 5A) was digested with EcoRI and resulting fragments were cloned into pUC119 for further characterization.

[0106] The insert containing plasmid was purified using the QIAGEN plasmid kit and sequenced. Nucleotide sequencing reaction was performed using the di-deoxy termination method, and was carried out with an ABI 310 sequencer. The exons and introns were determined by primer walking on both strands, and the size of the introns was estimated by sequencing in combination with agarose gel

electrophoresis. There appear to be only 3 exons coding for the C5-epimerase, with the longest exon coding for more than 50% of the protein. The exon-intron junctions (splice sites) precisely follow the gt-ag consensus rule. Based on the presence of introns and the precise match between the exons and the cDNA sequence, we believe that the genomic clone identified represents the functional gene of the C5-epimerase.

Cloning of the mouse C5-epimerase cDNA

[0107] One pair of primers was designed and based on the nucleotide sequence obtained by sequencing the exons of the genomic clone. The sense primer corresponds to bp 1-26 of the mouse ORF, starting from initiation codon ATG. The antisense primer corresponds to bp 1829-1854 without including the stop codon. PCR was performed by using a mouse liver QUICK-Clone™ cDNA (Clontech) as template at the conditions: 1 cycle of 94°C for 1 min, 30 cycles each of 94°C for 30 s, 60°C for 45 s and 72°C for 1 min, and a final extension at 72°C for 10 min. A strong band of about 2 kb was obtained, which was cloned into a TOPO™-TA Cloning vector (Invitrogen) and amplified and subsequently sequenced. By double strand sequencing it was found the mouse C5-epimerase clone is 1875 bp long, with a strong hydrophobic domain at N-terminal of the deduced peptide.

Northern blot analysis

[0108] The mouse multi-tissue mRNA blot was purchased from Clontech. The DNA probe from bovine cDNA clone was labeled with [α^{32} P]dCTP by Klenow enzyme from Boehringer Mannheim. The hybridization was carried out in ExpressHyb (Clontech) at 60°C for one hour and washed at high stringency. The membrane was exposed to Kodak film at -70°C overnight. The C5-epimerase enzyme is expressed in all tissues examined and the transcript is around 5 kb. It

seems that the liver has the highest expression for the transcript, while the spleen expresses a relatively lower level relative to β-actin in the same membrane.

Southern blot analysis

[0109] Southern blot analysis was performed according to Sambrook *et al.* (Sambrook *et al.*, 1989). Mouse genomic DNA was prepared with an Easy Prep kit (Pharmacia Biotech). 20 μg of genomic DNA was digested with restriction enzyme SacI, and separated on a 0.8% agarose gel by electrophoresis. After electrophoresis, the gel was treated with 0.1N NaOH for 30 min and neutralized in Tris-HCl buffer. The DNA fragments were transferred onto a nylon membrane. A 837 bp fragment of bovine C5-epimerase cDNA was labeled with [α^{32}P]dCTP by Klenow enzyme from Boehringer Mannheim and used as probe. The hybridization conditions were carried out as described for Northern analysis. The exposure time was 3 days.

[0110] To determine how many genes may potentially code for C5-epimerase, twenty micrograms of mouse genomic DNA purified from mouse liver was digested with restriction enzymes of ApaI, BamHI, EcoRI, EcoRV, HindIII, NcoI and XbaI respectively and separated on a 0.8% agarose gel by electrophoresis. The DNA separated in the gel was transferred to a Nylon membrane and was subsequently hybridized with a DNA probe from bovine coding sequence (1407 bp). The restriction map of the C5-epimerase genomic DNA suggests that there is only one gene coding for the C5-epimerase enzyme in mice.

Enzyme activity analysis

[0111] The activity of C5-epimerase was assessed according to the protocol as disclosed in Malmström *et al., J. Biol. Chem. 255*:3878-3883 (1980), which is herein incorporated by reference. Briefly, a mouse, transplanted with mastocytoma cells intramuscularly, was euthanized by cervical dislocation and

then dissected. The respective tissues, including the xenograft, were taken and were immediately homogenized in a buffer of 50 mM HEPES containing 100 mM KCl, 15 mM EDTA, 1% Triton X-100 and protease inhibitors. The homogenates were shaken at 4°C for 30 min and centrifuged. The supernatant was collected. Total protein concentration was determined by QuantiGold assay, and the specific activity of C5-epimerase was analyzed based on the release of ^3H (recovered as ^3H_2O) from a substrate polysaccharide according to the procedure described by Li *et al.* (Li *et al.* 1997). The C5-substrate used in the specific activity test is analyzed at least once a month by measuring only 50 µl C5-substrate working solution without any enzyme.

[0112] If the initial activity of the sample is >2000 cpm/50 µl, this is an indication that the sample is saturated and needs to be diluted. Dilution factors depend on the samples used, the saturation of the samples and on the salt concentration. The sample must contain not more than 50 mM salt (NaCl or KCl), because the C5-epimerase activity is partially or completely inhibited at higher salt concentrations.

[0113] Positive and negative controls are used in C5-epimerase activity assay. The positive control has to be standardized every two months to be sure that the stability has been preserved. Only a vector produced in the same cells as the sample can be used as a negative control. For example, for the C5-samples produced by baculovirus/insect cell expression system acetylcholinesterase produced with the same system has been used as a negative control.

[0114] During the prewarming of the C5 substrate solution, the samples were diluted if needed. 50 µl sample (enzyme) was added to the prewarmed substrate and incubated exactly for 1 h at + 37°C. After incubation 100 µl of a stop solution of enzyme reaction was added to the substrate-enzyme mixture, and this reaction mixture was transferred to a Wallac's 20 ml scintillation vial. 13 ml of epimerase assay scintillation cocktail was added to the vials and vortexed for 10 s. Radioactivity was measured in triplicate with a Wallac 1415 Liquid Scintillation Counter for 2 minutes each, after overnight incubation. The

scintillation counter gives the results as cpm/reaction volume (50 µl). If a sample has been diluted, the activity of the dilution buffer should be subtracted from sample's activity. In any case, the activity of the blank was subtracted from the activity of the sample before analyzing the results.

[0115] Specific activity was measured by dividing total activity (cpm/µl) by total protein concentration (mg/ml). Total protein concentration was measured by QuantiGold assay according to Stoschek, C.M., *Anal. Biochem. 160*:301-305 (1987), which is herein incorporated by reference. The unit of specific activity is cpm/mg/h, where h (hour) describes the time of the enzyme reaction.

EXAMPLE 2

Identification of the true N-terminus of C5-epimerase from coding sequence analysis of mouse gene, and expression of cloned cDNA.

[0116] Based on cloning and preliminary sequence analysis of the putative mouse C5-epimerase gene identified in Example 1, and based on alignment to the previously published bovine cDNA sequence, additional murine 5′-flanking DNA sequence was isolated, and a cDNA was cloned that contained this 5′-flanking DNA sequence.

[0117] To determine if this 5′-flanking DNA sequence might encode additional N-terminal peptide sequences that would represent the true N-terminus of the C5-epimerase encoded by the mouse gene, the mouse sequence (bold text in the compiled nucleotide sequence shown in Figure 1) was added to the bovine cDNA sequence, which was already in a computer file, using the bovine sequence and starting from point of greatest conservation (>96% amino acid identity). Then, the Gene Inspector program (Textco, USA) was used to search for open reading frames (ORFs) which are potential polypeptide coding sequences) in the compiled sequence. The result of the sequence alignment is shown in Figure 1 and the ORF analysis yielded the results shown in Figure 2.

[0118] In Figure 1, the fusion site between the new mouse sequence and the bovine sequence is indicated by the double colon "::". The sequence beginning

following the double colon is the bovine cDNA sequence. The sequence (in bold) 5′ to the fusion site is the additional murine 5′-flanking DNA sequence isolated as described above. The underlined sequence is the open reading frame that was found, showing the polypeptide coding sequence.

[0119] It is known that the native C5-epimerase enzyme is localized to the membranous golgi "compartment" (microsomal fraction) of cells (from liver). Therefore, the native mouse sequence should contain a suitable N-terminal signal for translocation to this compartment. To analyze this, the algorithm (program) of Nielsen *et al.*, *Protein Engineering 10*:1-6 (1997), was used. The algorithm analyzed the "signal" potential of the first 40-60 amino acids from each of the above polypeptide sequences. The same program was used to test the first 40 residues of the mouse syndecan-1 polypeptide sequence, as this is known to contain a secretion signal, as a sort of control for efficacy of the program, and the program positively identified this (data not shown). The analysis demonstrated strong signal potential for the first 33 residues.

[0120] Besides the 33 amino acid signal sequence already mentioned, the 154 additional N-terminal residues include additional cysteine residues which might form disulfide bonds and stabilize protein folding, and a predicted amidation site (residues 118-121) that might be relevant to posttranslational proteolytic processing. Further analyses of the complete sequence for C5-epimerase predicts hydrophobic stretches of polypeptide which could be buried, or traverse membrane(s).

[0121] Alignment analysis to other sequences found in databases reveal hotspots of homology. These results are summarized in Figures 4 and 5.

[0122] Figure 5 is a diagrammatic representation of the mouse C5-epimerase polypeptide sequence. As shown on Figure 5, the greatest evolutionary conservation ("hotspots" of homology) of sequence has occurred in the more C-terminal portion, in a highly hydrophobic stretch between amino acid residues Trp497 and Leu523, predicted to be buried in the protein's folded structure or traversing a membrane, possibly into the lumen of the golgi, where the enzyme

is known to act. The other most significant and extended stretch ("hotspot") of conservation occurs between residues Leu546 and His580, and might contain or comprise the active site of the enzyme. The functional significance of polypeptide sequence conservation (identity) ≥22% has been established by published studies of other proteins of known function (Branden, C., and Tooze, J., *Introduction to Protein Structure*, Garland Publishing, NY and London, pp. 100-101 (1991))., and Wilson, Kreychman, and Gerstein and the other authors cited therein, in "Quantifying the relations between protein sequence, structure, and function through traditional and probabilistic scores," available at bioinfo.mbb.yale.edu/e-print/ann-xfer-jmb/preprint. As explained in this article, precise function does not appear to be conserved below 30-40% sequence identity, whereas functional class is conserved for sequence identities as low as 20-25%. Below 20%, general similarity is no longer conserved.). At present, SWISS-MODEL will generate models for sequences which respond to these criteria: BLAST search P value: <0.00001; Global degree of sequence identity (SIM): >25% and Minimal projected model length - 25 amino acids.

[0123] Based on this, it is seen that the Drosophila sequence is more closely related (46.6%) than the *C. elegans* sequence (39.6%) to the mouse sequence.

[0124] In another type of sequence analysis, the predicted three-dimensional (3D) structure of the mouse C5-epimerase sequence was "threaded" against the 3D structures of Kelley, L.A., *et al.*, *Mol. Biol. 299*(2):499-520 (2000). This comparison indicated that the C5-epimerase sequence has a significant relationship to a chondroitinase (chondroitin AC/alginate lyase) domain, which is an alpha/alpha toroid. The chondroitin AC lyase is representative of a family of glycosaminoglycan degrading enzymes, and structure/function relationships have been elucidated from crystallography (Fethiere *et al., J. Mol. Biol. 288*:635-647 (1999). Remarkably, the most significant 3D similarity to the chondroitinase sequence was found to extend from Ala408, near the C-terminal end of an internal hydrophobic (transmembrane) stretch, to the C-terminus of the mouse C5-epimerase sequence, and that this stretch contains most of the conserved

sequence conservation likely indicates that it is a domain containing the active site.

[125] Based on all the above sequence analyses results, new recombinant C5-epimerase constructs were made, in addition to the first active tagged recombinant (bovine) C5-epimerase construct, for heterologous secretion-expression from baculovirus and InsectSelect (Invitrogen, USA) expression systems. The products from cloned insect cell lines so far characterized are summarized in Figure 6A. Four constructs are shown. The first construct is the tagged recombinant bovine C5-epimerase. The second construct is the tagged full length mouse C5-epimerase. The third construct is a tagged, chimeric construct between the mouse and bovine C5-epimerase sequences. The fourth construct is a tagged, truncated mouse sequence.

[126] In each of the recombinant constructs, the C5-epimerase was tagged. When tagged, the C5-epimerase sequence was preceded by a sequence as shown in Figure 6B which contains the EGT signal peptide linked to the EGT signal cleavage, an enterokinase cleavage site, six histidines, and finally the rTEV protease site. The EGT signal is from a protein of baculovirus (which infects insect cells). The FLAG sequence is an epitope-tag used for detecting and purifying recombinant protein according to the manufacturer's suggested protocol (Sigma) (Hopp, T. *et al.*, *Biotechnology 6*:1204-1210 (1988)). Enterokinase is an enzyme used to cleave off the sequence preceding its recognition site. The six consecutive histidines are another tag. The rTEV (recombinant tobacco edge virus) protease-site was also used to remove the preceding sequences. The EGT signal and FLAG™-tag (IBI) were obtained from constructs made in a modified pFastBac™ (Life Technologies) vector provided by Dr. Christian Over-Blom, VTT Biotechnology, P.O. Box 1500, FIN-02044, VTT, FINLAND. The purification of all recombinant proteins described in this application was FLAG-tag-based.

[127] The representative data from activity assays and protein analyses of these tagged recombinant C5-epimerases are shown in Figure 7, and Table I and Table

II. Figure 7 shows activity assay results of mouse C5-epimerase (mC5) that had been purified over anti-FLAG M1 according to the manufacturer's suggested protocol.

[128] The C5-epimerase activity assay to measure the activity of the heterologous protein was performed as in Example 1 above. Briefly, total protein was extracted from cultures transformed with each of the recombinant C5-epimerase constructs that were individually inoculated into insect cells using the InsectSelect expression systems. After the cells reached confluence, they were harvested and lysed and total protein was isolated and quantitated. C5-epimerase activity was measured as ^{3}H release from the epimerase substrate in a scintillation counter. Epimerase activity was measured against total protein. Figure 7 shows the activity with increasing volume of sample (diluted 1:2000). The total activity was 6360 cpm/μl. Protein analysis (using QuantiGold, Diversified Biotech) was analyzed according to Stoschek, C.M., *Anal. Biochem. 160*:301-305 (1987) and indicated that the concentration of protein was 3.2 μg/ml. Therefore, the specific activity was 2.0×10^{9} cpm/mg/h.

[129] Figure 8 shows a Western blot stained with anti-FLAG. Lane 1 contains molecular weight standards (New England Biolabs, Broad Range, prestained). Line 2 contains the full-length mouse C5-epimerase. The tagged full-length mouse C5-epimerase (that contains the N-terminal additional sequences found herein) has a length of 618 amino acids, a molecular weight (daltons) of 71189.1, an isoelectric point (pI) of 8.25 and a net charge at pH 7 of +4.01.

[130] Figure 9 is a Western blot of the culture medium taken from stable insect cell lines of the different clones for the four tagged recombinant C5-epimerases described above, stained with anti-FLAG antibody (020300). Lane 1 contains molecular weight standards as in Figure 8, with the molecular weights noted on the side of the gel. Lane 2 contains the truncated mouse C5-epimerase. Lane 3 contains the original bovine C5-epimerase. Lane 4 contains the mouse:bovine chimeric C5-epimerase in which the N-terminal mouse sequences are fused in frame to the bovine sequences, as shown in Figures 2 and 3. Lane 5 contains the

full-length mouse C5-epimerase. It can be seen that the chimeric mouse: bovine construct is approximately the same size as that of the full-length mouse construct.

[131] The relative activity of the different recombinant constructs was calculated based on the activity assays and densitometric analysis of the Western Blot and is shown in Table I, below. "TruncC5" is the shortened mouse C5-epimerase amino acid sequence where the first 154 amino acids have been removed such that the "TruncC5" sequence has the same N-terminus as the recombinant bovine sequence. "ExtC5" is the recombinant bovine C5-epimerase polypeptide, while "chC5" refers to the mouse:bovine chimeric C5-epimerase construct encoded by the nucleic acid sequence as shown in Figure 1. "mC5" refers to the full-length mouse C5-epimerase sequence.

Table I. Relative activities of different recombinant C5-epimerases,

Sample	Density	Sample (µl)	Density/ µl	Activity (cpm/µl)	Activity/Density (Cpm/densitometric unit)
truncC5	15984	12	1332.0	20	0.015
extC5	6451	12	537.6	7	0.013
chC5	14960	12	1246.7	3455	2.771
mC5	13804	12	1150.3	3681	3.200

[132] The Specific activities of the different partially-purified recombinant C5-epimerases is shown in Table II.

Table II. Specific activities of the Partially-purified recombinant C5-epimerases,

Sample	Total activity (cpm/µl)	Linearity (R^2)	Protein (mg/ml)	Specific Activity (cpm/mg/h)
truncC5	39.5	0.9905	0.0129	3.06×10^6
extC5	9.1	0.9978	0.0092	9.89×10^5

Sample	Total activity (cpm/µl)	Linearity (R^2)	Protein (mg/ml)	Specific Activity (cpm/mg/h)
chC5	919.7	0.9964	0.0026	3.54×10^8
mC5	2019	0.9969	0.0042	4.81×10^8

[133] The chimeric mouse:bovine construct that was made contains amino acid residues 34-154 of the N-terminal sequence of the mouse polypeptide sequence, immediately following the EGT-FLAG-His-RTEV elements as shown in Figure 6B. However, that recombinant enzyme appeared to be predominantly retained in the cytosol, probably due to the signaling potential of the mouse sequence.

Conclusion

[134] The addition of an N-terminal fragment of polypeptide (Asp34 to Asp154) from the mouse gene sequence enhances the activity of recombinant C5-epimerase enzyme by orders of magnitude, even though this piece of sequence does not contain the greatest interspecies conservation. The possible effect of tags on activity of the first recombinant bovine construct has been addressed (by tag removal; data not shown), and might account for a minor factor of the difference, but not to the extent of the orders of magnitude differences in specific activities between longer and shorter forms of recombinant C5-epimerase. Untagged expression constructs and structure-function studies are currently underway to better define the basis and mechanism for controlling the activity of this very important recombinant enzyme.